Molecular binding in post-Kohn–Sham orbital-free DFT

Alex Borgoo,*,† James A. Green,‡ and David J. Tozer*,‡

Department of Chemistry, Centre for Theoretical and Computational Chemistry, University of Oslo, P.O. Box 1033, Blindern, Oslo N-0315, Norway, and Department of Chemistry, Durham University, South Road, Durham DH1 3LE, UK

E-mail: ajborgoo@gmail.com; d.j.tozer@durham.ac.uk

Abstract

Molecular binding in post-Kohn-Sham orbital-free DFT is investigated, using non-interacting kinetic energy functionals that satisfy the uniform electron gas condition and which are inhomogeneous under density scaling. A parameter is introduced that quantifies binding and a series of functionals are determined from fits to near-exact effective homogeneities and/or Kohn-Sham non-interacting kinetic energies. These are then used to investigate the relationship between binding and the accuracy of the effective homogeneity and non-interacting kinetic energy at the equilibrium geometry. For a series of 11 molecules, the binding broadly improves as the effective homogeneity improves, although the extent to which it improves is dependent on the accuracy of the non-interacting kinetic energy; optimal binding appears to require both to be accurate simultaneously. The use of a Thomas-Fermi-von Weizsäcker form, augmented with a second gradient correction, goes some way towards achieving this, exhibiting molecular

*To whom correspondence should be addressed
†Oslo University
‡Durham University

binding on average. The findings are discussed in terms of the non-interacting kinetic potential and the Hellmann-Feynman theorem. The extent to which the functionals can reproduce the system-dependence of the near-exact effective homogeneity is quantified and potential energy curves are presented for selected molecules. The study provides impetus for including density scaling homogeneity considerations in the design of non-interacting kinetic energy functionals.

Introduction

The non-interacting kinetic energy, T_s, is a key component of the electronic energy in density functional theory (DFT). In regular Kohn–Sham calculations, T_s is evaluated exactly using the Kohn–Sham orbitals,[1]

$$T_\mathrm{s}[\{\phi_i\}] = -\frac{1}{2}\sum_i \int \phi_i(\mathbf{r})\nabla^2\phi_i(\mathbf{r})d\mathbf{r}, \tag{1}$$

leaving the exchange-correlation energy and potential to be approximated. Unfortunately, the solution of the Kohn–Sham equations that yields the orbitals has a computational cost that scales formally with the cube of system size, limiting the size of system that can be studied. If T_s could instead be accurately approximated as an explicit functional of the electron density ρ, then the need for the orbitals could be eliminated, reducing the computational cost and expanding the size of system that can be studied. It is therefore highly desirable to develop accurate non-interacting kinetic energy density functionals, $T_\mathrm{s}[\rho]$, for use in so-called orbital-free DFT, and this has been the subject of much research effort.[2–22] Such functionals also play a key role in frozen density embedding approaches.[23–25]

Recently, we have been investigating[26–28] the scaling behaviour of $T_\mathrm{s}[\rho]$, in the hope that it may provide information that can aid the development of improved functionals. The most common type of scaling is *coordinate scaling*.[29] It is well-known that the exact $T_\mathrm{s}[\rho]$ is

homogeneous of degree 2 under coordinate scaling, meaning that it satisfies

$$T_{\rm s}[\rho_\lambda] = \lambda^2 T_{\rm s}[\rho], \tag{2}$$

where the coordinate scaled density is

$$\rho_\lambda(\mathbf{r}) = \lambda^3 \rho(\lambda \mathbf{r}). \tag{3}$$

All functionals in the present study satisfy Eqn. (2) and this type of scaling will not be discussed further.

Our research has instead focussed on a less-well-known form of scaling, termed *density scaling* (or more precisely, ensemble density scaling;[28] see also Ref. 30). A functional $T_{\rm s}[\rho]$ is homogeneous of degree k under density scaling if it satisfies

$$T_{\rm s}[\xi\rho] = \xi^k T_{\rm s}[\rho], \tag{4}$$

or equivalently (for $k \neq 0$),[31]

$$k = \frac{\int \rho(\mathbf{r}) \frac{\delta T_{\rm s}[\rho]}{\delta \rho(\mathbf{r})} d\mathbf{r}}{T_{\rm s}[\rho]}, \tag{5}$$

where $\frac{\delta T_{\rm s}[\rho]}{\delta \rho(\mathbf{r})}$ is the non-interacting kinetic potential. Evaluation of the quantity k using Eqn. (5) therefore provides a simple mechanism for quantifying the behaviour of any functional $T_{\rm s}[\rho]$ under density scaling. If the value of k is system-independent then the functional is homogeneous of degree k. If the value of k is system-dependent, then the functional is inhomogeneous and in such cases k is termed an 'effective homogeneity'.[32] The degree of system-dependence provides a measure of the degree of inhomogeneity.

The exact $T_{\rm s}[\rho]$ is inhomogeneous under density scaling as can be readily seen by evaluating Eqn. (5) for two disparate systems: the uniform electron gas is described exactly by Thomas-Fermi theory,[2,3] for which $k = 5/3$, whereas one-electron systems are described exactly by the von Weizsäcker functional form,[33] for which $k = 1$. In a recent study,[26] we

further quantified the system-dependence of k for the exact $T_{\rm s}[\rho]$ by evaluating Eqn. (5) using experimental and near-exact calculated data for a series of atoms and small molecules at equilibrium geometries. We concluded that the influence of the integer discontinuity was small and that the values of k associated with the potential that averages over the integer discontinuity – which is most appropriate for the development of a continuum functional such as a generalised gradient approximation (GGA) functional – did not exhibit significant system dependence for systems with more than a few electrons; the functional is only mildly inhomogeneous for the systems considered. The average near-exact effective homogeneity was $k = 1.55$, notably lower than the Thomas-Fermi value of $k = 5/3$.

In a subsequent paper,[27] we investigated the influence of imposing this average near-exact value, paying particular attention to the degree of molecular binding in three small molecules in a post-Kohn–Sham approach. Binding is a key consideration in orbital-free DFT, since it is absent in Thomas-Fermi theory[34] and is also absent in most (though not all) generalised gradient approximation (GGA) forms. We considered a single-term functional

$$T_{\rm s}[\rho] = cT'_{\rm s}[\rho] \tag{6}$$

where

$$T'_{\rm s}[\rho] = \int \rho^{5/3}(\mathbf{r}) x^n(\mathbf{r}) d\mathbf{r}. \tag{7}$$

Here, $x = \frac{|\nabla\rho|}{\rho^{4/3}}$ is the usual dimensionless quantity and c and n are parameters. The functional in Eqn. (6) is homogeneous of degree $k = \frac{1}{3}(5-n)$ for all values of c. A value of n was therefore chosen and the value of c was subsequently determined through a linear regression to Kohn–Sham $T_{\rm s}[\{\phi_i\}]$ values for a series of atoms and molecules at equilibrium geometries. We observed that when $n = 0$, i.e. the functional was homogeneous of degree $k = 5/3$, it did not bind any of the three molecules, consistent with Thomas-Fermi theory. By contrast, when $n = 0.343$, i.e. the functional was homogeneous of degree $k = 1.55$, all three molecules were bound, albeit without quantitative accuracy. Furthermore, when the power n was

chosen to optimise the values of T_s – rather than the value of k – two of the molecules failed to bind. See Ref. 27 for full details. The key conclusion of the study was that forcing k to be near-exact may be beneficial for molecular binding.

The purpose of the present study is to develop the approach introduced in Ref. 27 and to provide a thorough investigation of molecular binding in post-Kohn–Sham orbital-free DFT. We have two specific aims. The first aim is to improve the functional form and behaviour under density scaling. The form in Eqn. (6) used in Ref. 27 has limited flexibility; it does not satisfy the uniform electron gas (UEG) condition; and it is homogenous under density scaling, whereas we know that the exact $T_s[\rho]$ is inhomogeneous. We shall therefore use a flexible form that satisfies the UEG and is inhomogeneous, meaning that the k values associated with the functional are system-dependent. A key question is how well the system-dependence of the near-exact k can be reproduced. The second aim is to investigate the functional characteristics that affect binding. The results of Ref. 27 suggest that forcing k to be near-exact may be beneficial, but it is important to note that the optimal binding was obtained using a functional where the value of T_s at the equilibrium geometry was *also* optimised, through the choice of the prefactor c in Eqn. (6). We shall investigate the role of both k and T_s in binding. We commence by describing our methodology. The results are then discussed and conclusions are drawn.

Methodology

Computational details

Following Ref. 27 and other studies,[19,35–37] all calculations are performed in a post-Kohn–Sham manner, using densities/orbitals determined using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional[38] and the aug-cc-pVTZ basis set.[39–41] These densities approximate those that would be obtained from a solution of the Euler equation using the exact $T_s[\rho]$. The total electronic energy associated with an approximate functional T_s^{approx} is

obtained as

$$E^{\mathrm{approx}} = E^{\mathrm{PBE}} - T_{\mathrm{s}}[\{\phi_i^{\mathrm{PBE}}\}] + T_{\mathrm{s}}^{\mathrm{approx}}[\rho^{\mathrm{PBE}}] \tag{8}$$

where a superscript 'PBE' indicates a PBE quantity computed from a regular Kohn–Sham calculation. In all cases, the accuracy of $T_{\mathrm{s}}^{\mathrm{approx}}$ and E^{approx} are quantified by comparing with the Kohn-Sham values. For molecules, all the T_{s} and k values used in the functional development and presented in the Figures were determined at the equilibrium geometries of Ref. 26. All calculations use a spin-restricted formalism.

In order to investigate the factors that affect binding, it is helpful to define a parameter that quantifies the degree of binding in a given molecule. If our concern is the energy of the molecule relative to that of completely separated atoms, then the atomisation energy would be the appropriate parameter. However, the incorrect dissociation arising from errors in the PBE description of exchange-correlation would necessitate an unrestricted spin-polarised formalism and this is beyond the scope of the present study, since our near-exact k for open-shell systems were determined in a spin-restricted formalism. In practical calculations where one is computing a molecular geometry, it can be sufficient to ask whether the energy of the molecule reduces in the vicinity of the experimental geometry, without reference to the energy of completely separated atoms. For diatomic molecules, this can be naturally quantified using the following parameter,

$$b = \left(\frac{E^{\mathrm{approx}}(2r_{\mathrm{e}}) - E^{\mathrm{approx}}(r_{\mathrm{e}})}{E^{\mathrm{PBE}}(2r_{\mathrm{e}}) - E^{\mathrm{PBE}}(r_{\mathrm{e}})} \right) \times 100 \tag{9}$$

where r_{e} is the experimental bond length. (Similar results would be obtained if the PBE r_{e} value were to be used, but we choose to use the experimental value since this is the geometry at which our near-exact k were computed). For polyatomic molecules, whose symmetry is such that there is only one bond length, the same expression can again be used, providing the other internal coordinates are kept fixed throughout (i.e. it provides a measure of binding

along one chosen pure bond stretching coordinate). Throughout this study, we shall therefore use b as a measure of 'binding' in our calculations using approximate functionals. A positive value of b signifies that the approximate functional correctly predicts a decrease in energy (i.e. binding) in moving from $2r_{\mathrm{e}}$ to r_{e}, whereas a negative value signifies an incorrect increase (i.e. repulsion). The optimal value is $b = +100\%$.

Our investigation of binding will focus largely on the diatomics Cl_2, F_2, CO, N_2, HCl, HF, and the polyatomics H_2S, H_2O, PH_3, NH_3 and CH_4, for which near-exact k values are available in the supplementary material of Ref. 26. We define the average binding parameter for these 11 molecules as

$$B = \frac{1}{11}\sum_{i=1}^{11} b_i, \tag{10}$$

where b_i is the value of b for system i.

Approximate $T_{\mathrm{s}}[\rho]$ functionals

We consider a flexible generalised gradient approximation (GGA) functional form,

$$T_{\mathrm{s}}[\rho] = T_{\mathrm{s}}^{\mathrm{TF}}[\rho] + c_1 T_{\mathrm{s}}^{\mathrm{W}}[\rho] + c_2 T_{\mathrm{s}}'[\rho] \tag{11}$$

where

$$T_{\mathrm{s}}^{\mathrm{TF}}[\rho] = C_{\mathrm{TF}} \int \rho^{5/3}(\mathbf{r}) d\mathbf{r} \tag{12}$$

is the regular Thomas-Fermi functional with $C_{\mathrm{TF}} = \frac{3}{10}(3\pi^2)^{2/3}$, which is homogeneous of degree $k = 5/3$. The functional

$$T_{\mathrm{W}}[\rho] = \frac{1}{8} \int \rho^{5/3}(\mathbf{r}) x^2(\mathbf{r}) d\mathbf{r} \tag{13}$$

is the regular von Weizsäcker functional, which is homogeneous of degree $k = 1$ (and is the second-order correction in the gradient expansion). And $T_{\rm s}'[\rho]$ is the functional defined in Eqn. (7), which is homogeneous of degree $k = \frac{1}{3}(5-n)$. The first two components are, of course, special cases of $T_{\rm s}'[\rho]$. We choose to use the simple form in Eqn. (11) for several reasons. The fact that it comprises a linear combination of homogeneous functionals means that Eqn. (5) is easily evaluated, yielding an expression for k in terms of the energy components,

$$k = \frac{\frac{5}{3}T_{\rm s}^{\rm TF}[\rho] + c_1 T_{\rm s}^{\rm W}[\rho] + \frac{c_2}{3}(5-n)\,T_{\rm s}'[\rho]}{T_{\rm s}^{\rm TF}[\rho] + c_1 T_{\rm s}^{\rm W}[\rho] + c_2 T_{\rm s}'[\rho]} \tag{14}$$

where we have used the fact that the individual components are homogeneous of degree $k = 5/3$, 1, and $\frac{1}{3}(5-n)$, respectively. For $c_1 \neq 0$ or $c_2 \neq 0$, the value of k does not reduce to a constant and so system-dependence is naturally introduced − the functional is inhomogeneous, as required. Furthermore, given that the first two components in Eqn. (11) are homogeneous of degree 5/3 and 1, a linear combination of the two will yield a value of k that is intermediate between the two values, as required. The addition of the third term provides further flexibility. Finally, Eqn. (11) satisfies the UEG, because $x = 0$ for this system.

We consider two forms forms based on Eqn. (11). The first has $c_1 \neq 0; c_2 = 0$, and so represents a Thomas-Fermi-von Weizsäcker form; this form will hereafter be denoted the 'two-term' form, defined by a single parameter, c_1. The second is a more flexible form with $c_1 \neq 0; c_2 \neq 0$, which will hereafter be denoted the 'three-term' form, defined by three parameters, c_1, c_2, and n. A key consideration is how to determine these parameters. Given that we wish to quantify the influence of k and $T_{\rm s}$ on binding − and that both of these quantities depend on all the parameters defining the functional − it is natural to define the parameters as those that minimise

$$\Omega = \alpha\Omega_k + (1-\alpha)\Omega_{T_{\rm s}} \tag{15}$$

where Ω_k and Ω_{T_s} are the mean absolute percentage errors in k (determined using Eqn. (14)) and T_s (determined using Eqn. (11)), respectively, relative to the near-exact system-dependent k values (averaged over the integer discontinuity) from Ref. 26 and the Kohn–Sham $T_s[\{\phi_i\}]$, respectively, for a training set. (Note that one could alternatively define Ω_k relative to k values determined using PBE, rather than experimental/near-exact quantities; we have verified that this alternative approach yields essentially identical functionals and conclusions because the k values from the two approaches are very similar for all but the lightest atoms). Following Ref. 27, we define our training set to be He, Be, Ne, Mg, Ar, SO_2, Cl_2, F_2, CO, PH_3, H_2S, N_2, HCl, NH_3, HF, H_2O, and CH_4. The set comprises the 11 molecules used in Eqn. (10), plus 5 atoms and the additional SO_2 molecule, which was omitted from Eqn. (10) due to convergence problems at stretched bond lengths. By varying the parameter α from 0 to 1, a series of functionals can be determined whose emphasis shifts from optimal T_s, through a more balanced description of T_s and k, to optimal k. The minimisation of Ω was performed using the Mathematica program.[42]

Results and Discussion

We commenced by determining a series of two- and three-term functionals, each associated with an α value in the range $0 \leq \alpha \leq 1$, in steps of 0.01, and then quantifying the average binding over the 11 molecules. We define new quantities $\Omega_k^{\rm mol}$ and $\Omega_{T_s}^{\rm mol}$ to be the mean absolute percentage errors in k and T_s, respectively, relative to the near-exact/Kohn–Sham values, for the 11 molecules. Figures 1(a) and 1(b) present the values of $\Omega_k^{\rm mol}$ and $\Omega_{T_s}^{\rm mol}$, respectively, determined using each functional, as a function of α. Figure 1(c) presents the average binding, B in Eqn. (10), determined using each functional, as a function of α. Each point corresponds to an individual optimisation; the raw data is presented, rather than smoothly interpolated curves.

First consider the two-term functionals. As α increases from 0 to about 0.7, there is

little variation in the three quantities. The value of Ω_k^{mol} is approximately 4% whilst $\Omega_{T_\text{s}}^{\text{mol}}$ is approximately 0.1%. The value of the average binding parameter B is close to -50%, indicating strongly repulsive interactions. As α increases beyond about 0.7 (i.e. as the fitting procedure increasingly emphasises k over T_s) the k values improve, with Ω_k^{mol} reducing to approximately 0.6%, although this is obtained at the expense of a less accurate T_s, with $\Omega_{T_\text{s}}^{\text{mol}}$ increasing to over 10%. The average binding parameter B increases, to approximately -22%, indicating a reduction in repulsion, on average.

Next consider the more flexible, three-term functionals. At small α, the results are very similar to those of the two-term functionals. As α increases beyond about 0.2, however, the quantity Ω_k^{mol} decreases notably to approximately 0.6% , with only a *minimal* increase in $\Omega_{T_\text{s}}^{\text{mol}}$ to approximately 1%, and this behaviour is maintained up to $\alpha = 0.99$. This is associated with an increase in the average binding parameter B to approximately $+12\%$. The positive value indicates binding on average, albeit significantly underestimated. Interestingly, when $\alpha = 1$, i.e. when the fit is purely to k, with no T_s information used, both $\Omega_{T_\text{s}}^{\text{mol}}$ and B jump abruptly to the two-term value, with no discernible change in Ω_k^{mol}. Examination of the functional parameters indicates that the three-term functional has essentially collapsed to the two-term functional. A marginal improvement in k can therefore be achieved (to the significant detriment of T_s) by eliminating the third term in the functional.

Overall, Figure 1 suggests that, on average, improving k at the equilibrium geometry does improve binding, although the extent that the binding improves is dependent on the accuracy of T_s at that geometry. For the two-term functional, the improvement in k is associated with a significant degradation in T_s and the binding improvement is modest. For the three-term functional, however, the improvement in k is associated with minimal degradation in T_s (except at $\alpha = 1$) and the improvement in binding is much more pronounced. Optimal binding would therefore appear to require k and T_s to be accurate simultaneously, which is consistent with Ref. 27, where the functional that exhibited optimal binding was optimised against both k (via the exponent) and T_s value (via the prefactor) independently.

Insight into this observation may be obtained from a consideration of Eqn. (5). The numerator contains the non-interacting kinetic potential $\frac{\delta T_{\rm s}[\rho]}{\delta\rho(\mathbf{r})}$, which, in a rigorous Euler formulation of orbital-free DFT, is a key quantity in molecular binding – it governs the accuracy of the density and hence the nuclear forces and the shape of the potential energy curve via the Hellmann-Feynman theorem. An accurate numerator is not a sufficient condition to ensure an accurate kinetic potential, but it is clearly desirable. If k is accurate but $T_{\rm s}$ is inaccurate then the numerator will also be inaccurate. Only when k and $T_{\rm s}$ are both accurate can we be sure that the numerator is accurate.

Having considered the average behaviour and dependence on α in Figure 1, we now go on to consider individual molecules for selected functionals. We consider the two- and three-term functionals evaluated using $\alpha = 0$, since these provide a benchmark for what happens when no homogeneity information is used in the fit; these functionals will hereafter be denoted 'two-term [$\alpha = 0$]' and 'three-term [$\alpha = 0$]'. We also consider the two functionals that exhibit the optimal (i.e. maximum) value of B in Figure 1(c), namely the two-term functional with $\alpha = 1$ and the three-term functional with $\alpha = 0.99$; these functionals will hereafter be denoted 'two-term [$\alpha = \alpha_{\rm opt}$]' and 'three-term [$\alpha = \alpha_{\rm opt}$]'. The parameters defining these four functionals are presented in Table 1. It is noteworthy that for the two-term [$\alpha = 0$] functional, the von Weizsäcker prefactor is $c_1 = 0.120$, close to the exact first-order gradient expansion of 1/9. The three-term [$\alpha = 0$] functional is a very similar functional with $c_1 = 0.115$ and a very small c_2 value of 0.006. Figure 2 plots the enhancement factor[22] $F_{\rm s}(x)$, defined by

$$T_{\rm s}[\rho] = \int \rho^{5/3} F_{\rm s}(x) d\mathbf{r}, \tag{16}$$

as a function of x, for the four functionals. The initial negative slope for the three-term [$\alpha = \alpha_{\rm opt}$] functional is a consequence of the negative c_2 coefficient. In light of the above discussion of the kinetic potential and Hellmann-Feynman theorem, we are presently investigating the potentials associated with these four functionals, comparing them with near-exact potentials and also establishing whether the small n values in the three-term functionals can lead to

any numerical issues. Results will be presented in a forthcoming publication.

Figure 3 presents a scatter plot of absolute percentage error in T_s vs absolute percentage error in k for the four functionals; each point corresponds to one of the 11 molecules. Consistent with Figure 1, the two-term $[\alpha = 0]$ and three-term $[\alpha = 0]$ functionals both yield high quality T_s, but relatively large errors in k. We note that for the $[\alpha = 0]$ functionals the T_s errors are generally slightly larger for three-term, rather than two-term, which seems counterintuitive, given the extra flexibility. However, this is simply a consequence of the fact that the functionals were defined by minimising Ω in Eqn. (15), which holds both atomic and molecular information. The minimised value of Ω at $\alpha = 0$ is marginally smaller for the three-term form, as expected. The two-term $[\alpha = \alpha_{\text{opt}}]$ functional yields significantly improved k at the expense of inaccurate T_s. Only the three-term $[\alpha = \alpha_{\text{opt}}]$ functional yields relatively good quality k and T_s, simultaneously.

Figure 4 presents the binding parameters of individual molecules, b, as a function of absolute percentage error in k. Again, each point corresponds to one of the 11 molecules. First consider the two-term $[\alpha = 0]$ and three-term $[\alpha = 0]$ results. The results from the two functionals are rather similar, with the binding broadly increasing as the error in k reduces. Indeed, we observe that a linear regression leads in both cases to an intercept of $b \approx 80\%$, suggesting that were the functionals able to reduce the error in k sufficiently, then the binding would be reasonably reproduced! Unfortunately, neither functional is able to reduce k sufficiently to test this. The results clearly illustrate our view that accurate binding requires accurate k and accurate T_s: both functionals yield accurate T_s throughout (see Figure 3) and so the determining factor is the accuracy of k.

Next consider the two-term $[\alpha = \alpha_{\text{opt}}]$ and three-term $[\alpha = \alpha_{\text{opt}}]$ results. Both functionals significantly reduce the error in k and exhibit a broad increase in binding as that error reduces, although the binding is uniformly underestimated, particularly for the two-term $[\alpha = \alpha_{\text{opt}}]$ functional. Again, this can be understood from a consideration of the accuracy of k and T_s: both functionals are able to reduce the error in k down to below 0.5% for most

of the systems, but the errors in T_s are large for the two-term functional and smaller, but still significant, for the three-term functional (see Figure 3). This is reflected in the binding, which improves from two-term to three-term, but remains underestimated. If the errors in T_s could be further reduced without degradation of k, then we would anticipate that the binding would be further improved. This provides a clear impetus for future functional development.

Another aim of the present study is to assess how well the system-dependence of the near-exact k can be reproduced by approximate functionals. Having identified four functionals, we can now quantify this. Figures 5(a) and 5(b) present k values for the $[\alpha = 0]$ and $[\alpha = \alpha_{\rm opt}]$ functionals, respectively, compared to the near-exact values of Ref. 26. Also shown are the Thomas-Fermi values of $k = 5/3$. The two- and three-term results are essentially indistinguishable. The $\alpha = 0$ functionals in Figure 5(a) yield k values that are smaller than 5/3, which is simply a consequence of adding the von Weizsäcker term (homogeneous of degree 1) to Thomas-Fermi (homogeneous of degree 5/3). The results are, however, significantly larger than the near-exact values. By contrast, the $[\alpha = \alpha_{\rm opt}]$ functionals in Figure 5(b) yield k values that are in much better agreement with the near-exact values. For the majority of the molecules, the k values are in excellent agreement with the near-exact values. The notable exceptions are H_2O, HF, and F_2, and it is these molecules that exhibit the smallest b values in Figure 4 for both functionals. Neither functional reproduces the system-dependence for the atoms.

The emphasis in this study has been on binding, as probed by the parameters b and B. We close by presenting full potential energy curves for representative systems, aligned at the longest bond length in each case. Figures 6 and 7 present curves for CO and F_2, respectively, which are both in the training set used to determine the functionals. Figure 8 presents curves for P_2, which is not in the fitting set. Parts (a) and (b) of each figure show results determined using the $[\alpha = 0]$ and $[\alpha = \alpha_{\rm opt}]$ functionals, respectively. Also shown are the curves from Thomas-Fermi theory and the Kohn–Sham curves.

The same observations are made for all three molecules. The two-term $[\alpha = 0]$ and three-

term [$\alpha = 0$] functionals yield virtually indistinguishable curves that are highly repulsive, though less repulsive than Thomas-Fermi. By contrast, the two-term [$\alpha = \alpha_{\mathrm{opt}}$] and three-term [$\alpha = \alpha_{\mathrm{opt}}$] functionals are distinguishable, with the three-term being closer to the Kohn–Sham curve in all cases. Of course, quantitative accuracy has not been achieved and we note that the [$\alpha = \alpha_{\mathrm{opt}}$] curves for CO in Figure 6(b) exhibit the same unphysical maxima at large r that was observed in Ref. 27. The results in Figures 6, 7, and 8 are fully consistent with the average and individual molecule results discussed above.

Conclusions

Molecular binding in post-Kohn-Sham orbital-free DFT has been investigated, using non-interacting kinetic energy functionals that satisfy the uniform electron gas condition and which are inhomogeneous under density scaling. A parameter was introduced that quantifies binding and a series of functionals were determined from fits to near-exact effective homogeneities and/or Kohn–Sham non-interacting kinetic energies. These were then used to investigate the relationship between binding and the accuracy of the effective homogeneity and non-interacting kinetic energy at the equilibrium geometry. For a series of 11 molecules, the binding broadly improves as the effective homogeneity improves, although the extent to which it improves is dependent on the accuracy of the non-interacting kinetic energy; optimal binding appears to require both to be accurate simultaneously. The use of a Thomas-Fermi-von Weizsäcker form, augmented with a second gradient correction, goes some way towards achieving this, exhibiting molecular binding on average. The findings were discussed in terms of the non-interacting kinetic potential and the Hellmann-Feynman theorem. The extent to which the functionals can reproduce the system-dependence of the near-exact effective homogeneity was quantified and potential energy curves were presented for selected molecules. The study provides impetus for including density scaling homogeneity considerations in the design of non-interacting kinetic energy functionals.

Acknowledgments

A.B. acknowledges support from the European Community Seventh Framework Programme (FP7, 2007-2013), Marie Curie Intra-European Fellowship, under the Grant Agreement No. FP7-254150 (OF-DFT/MCHF project). Furthermore A.B. grateful for support from the European Research Council advanced grant FP7 project ABACUS (grant number 267683).

References

(1) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(2) Fermi, E. *Rend. Accad. Lincei* **1927**, *6*, 602.

(3) Thomas, L. *Math. Proc. of the Cambridge Phil. Soc.* **1927**, *23*, 542–546.

(4) Smargiassi, E.; Madden, P. A. *Phys. Rev. B* **1994**, *49*, 5220–5226.

(5) Xia, J.; Huang, C.; Shin, I.; Carter, E. A. *J. Chem. Phys.* **2012**, *136*, 084102.

(6) Lee, H.; Lee, C.; Parr, R. G. *Phys. Rev. A* **1991**, *44*, 768–771.

(7) Lembarki, A.; Chermette, H. *Phys. Rev. A* **1994**, *50*, 5328–5331.

(8) Tran, F.; Wesołowski, T. A. *Int. J. Quant. Chem.* **2002**, *89*, 441–446.

(9) Karasiev, V.; Trickey, S.; Harris, F. *J. of Comp.-Aided Materials Design* **2006**, *13*, 111–129.

(10) Laricchia, S.; Fabiano, E.; Constantin, L. A.; Della Sala, F. *J. Chem. Theo. and Comp.* **2011**, *7*, 2439–2451.

(11) Constantin, L. A.; Fabiano, E.; Laricchia, S.; Della Sala, F. *Phys. Rev. Lett.* **2011**, *106*, 186406.

(12) Ernzerhof, M. *J. Mol. Struct.: THEOCHEM* **2000**, *501–502*, 59–64.

(13) Perdew, J. P. *Phys. Rev. A* **1992**, *165*, 79 – 82.

(14) Ou-Yang, H.; Levy, M. *Int. Jour. Quant. Chem.* **1991**, *40*, 379–388.

(15) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1986**, *33*, 8800–8802.

(16) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671–6687.

(17) Lacks, D. J.; Gordon, R. G. *J. Chem. Phys.* **1994**, *100*, 4446–4452.

(18) Thakkar, A. J. *Phys. Rev. A* **1992**, *46*, 6920–6924.

(19) Karasiev, V. V.; Jones, R. S.; Trickey, S. B.; Frank, E. H. In *New Developments in Quant. Chem.*; Paz, J. L., Hermández, A. J., Eds.; Transworld research network, 2009; Chapter 2, pp 25–54.

(20) Karasiev, V.; Trickey, S. *Computer Physics Communications* **2012**, *183*, 2519 – 2527.

(21) Chan, G. K.-L.; Cohen, A. J.; Handy, N. C. *J. Chem. Phys.* **2001**, *114*, 631–638.

(22) Laricchia, S.; Constantin, L. A.; Fabiano, E.; Della Sala, F. *J. Chem. Theo. and Comp.* **2014**, *10*, 164–179.

(23) Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1993**, *97*, 8050–8053.

(24) Wesolowksi, T. A. In *Comp. Chem.: Reviews of Current Trends;*; Leszczynski, J., Ed.; World Scientific: River Edge, 2006; Vol. 10; pp 1–82.

(25) Götz, A. W.; Beyhan, S. M.; Visscher, L. *J. Chem. Theo. and Comp.* **2009**, *5*, 3161–3174.

(26) Borgoo, A.; Teale, A. M.; Tozer, D. J. *J. Chem. Phys.* **2012**, *136*, 034101. Note that the average effective homogeneities for He and Be in Figure 4 are incorrect; the correct values have been used in the present study.

(27) Borgoo, A.; Tozer, D. J. *J. Chem. Theo. and Comp.* **2013**, *9*, 2250–2255.

(28) Borgoo, A.; Teale, A. M.; Tozer, D. J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14578–14583.

(29) Levy, M.; Perdew, J. P. *Phys. Rev. A* **1985**, *32*, 2010–2021.

(30) Fabiano, E.; Constantin, L. A. *Phys. Rev. A* **2013**, *87*, 012511.

(31) Liu, S.; Parr, R. G. *Phys. Rev. A* **1996**, *53*, 2211–2219.

(32) Zhao, Q.; Morrison, R. C.; Parr, R. G. *Phys. Rev. A* **1994**, *50*, 2138–2142.

(33) von Weizsäcker, C. *Z. Phys. A* **1935**, *96*, 431–458.

(34) Teller, E. *Rev. Mod. Phys.* **1962**, *34*, 627–631.

(35) Iyengar, S. S.; Ernzerhof, M.; Maximoff, S. N.; Scuseria, G. E. *Phys. Rev. A* **2001**, *63*, 052508.

(36) Constantin, L. A.; Ruzsinszky, A. *Phys. Rev. B* **2009**, *79*, 115117.

(37) Perdew, J. P.; Levy, M.; Painter, G. S.; Wei, S.; Lagowski, J. B. *Phys. Rev. B* **1988**, *37*, 838–843.

(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(39) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(40) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(41) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(42) *Mathematica Version 9.0*; Wolfram Research, Inc.: Champaign, Illinois, 2012.

Table 1: Parameters defining selected functionals; see Eqn. (11).

	two-term $[\alpha = 0]$	two-term $[\alpha = \alpha_{\rm opt}]$	three-term $[\alpha = 0]$	three-term $[\alpha = \alpha_{\rm opt}]$
c_1	0.119832	0.273776	0.115166	0.268960
c_2	-	-	0.006118	-0.230783
n	-	-	0.299001	0.298851

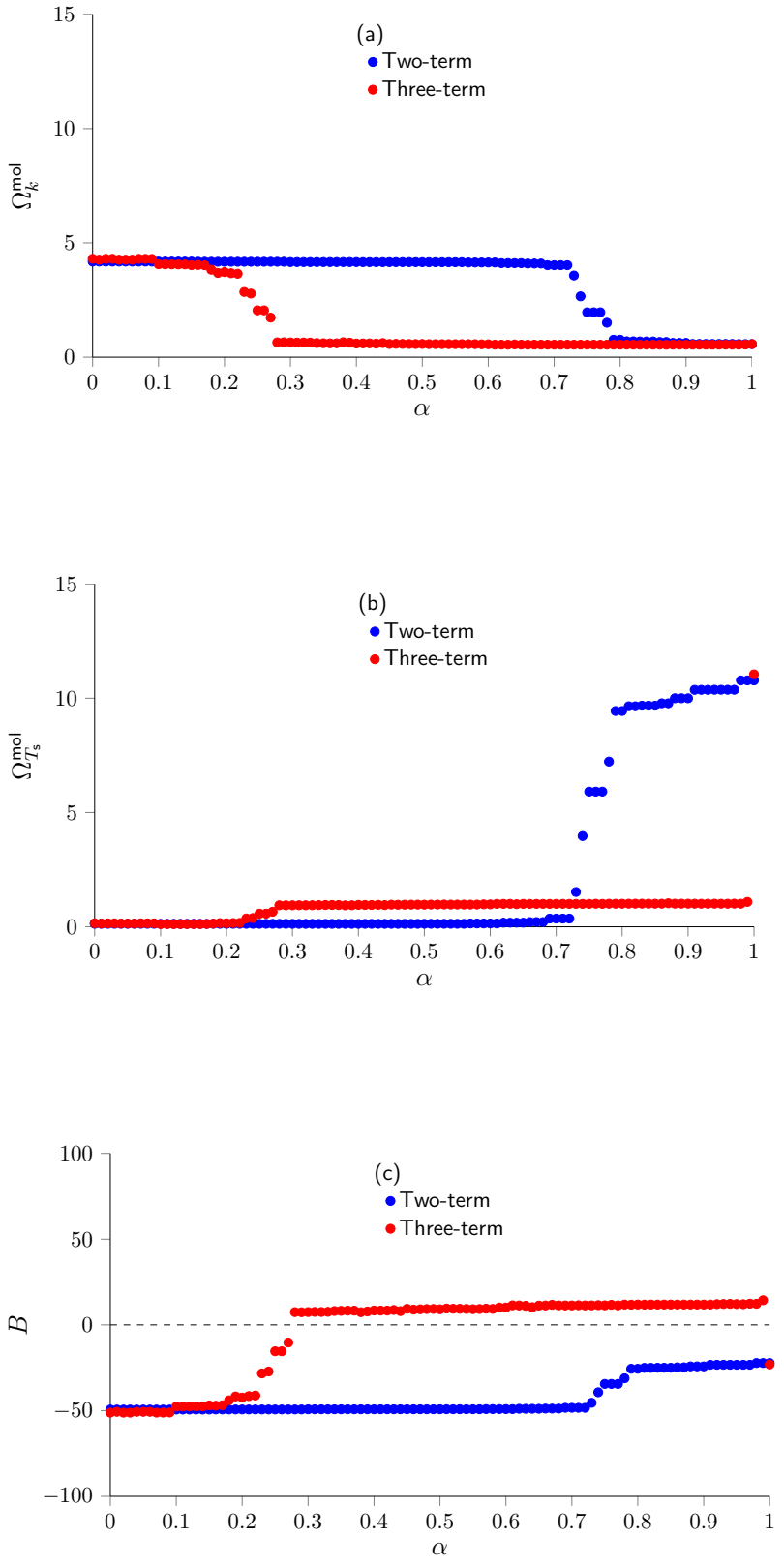

Figure 1: (a) Mean absolute percentage errors in k; (b) mean absolute percentage errors in T_{s}; and (c) average binding parameters B, for 11 molecules, plotted as a function of α in Eqn. (15)

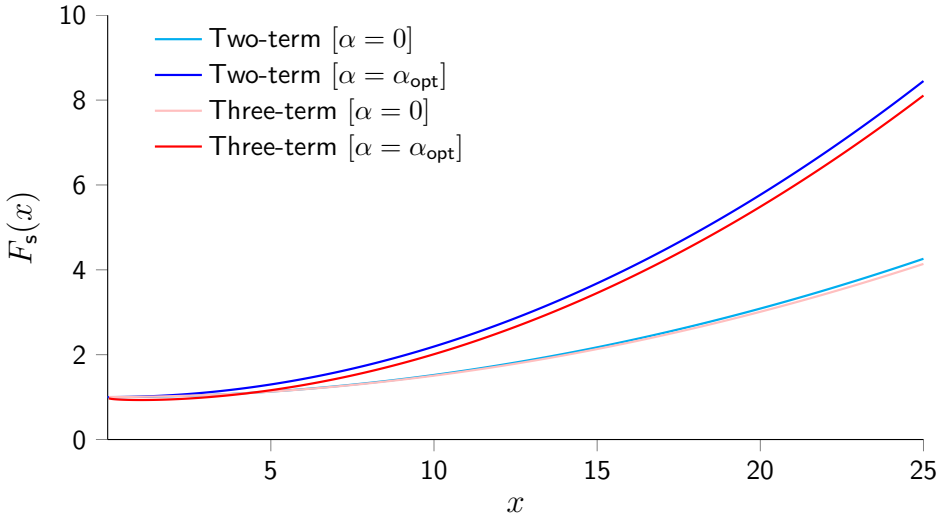

Figure 2: Enhancement factors for the four functionals in Table 1.

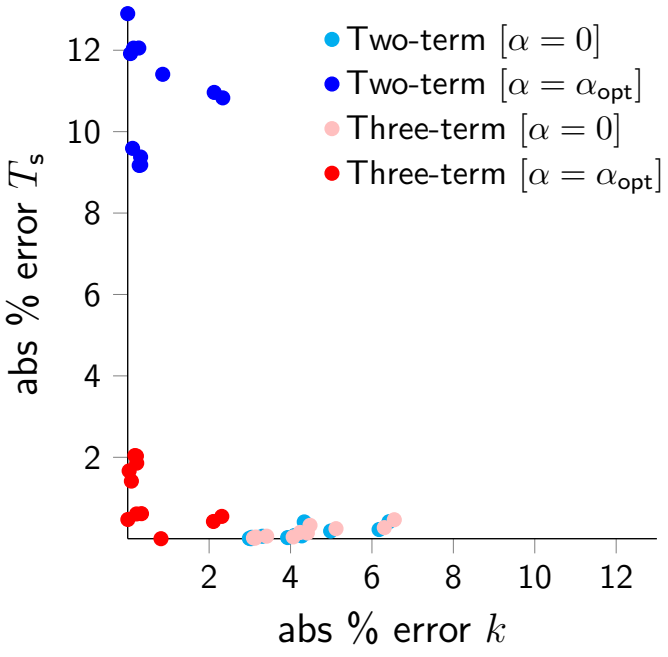

Figure 3: Scatter plot of absolute percentage errors in T_{s} and k, for 11 molecules.

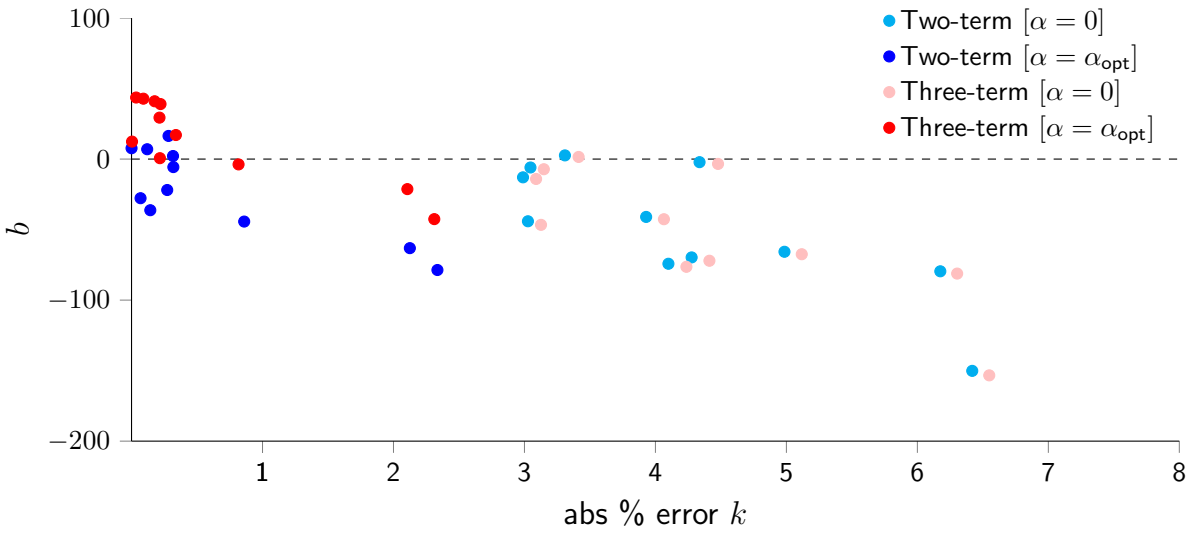

Figure 4: Individual molecular binding parameters, plotted as a function of absolute percentage errors in k, for 11 molecules.

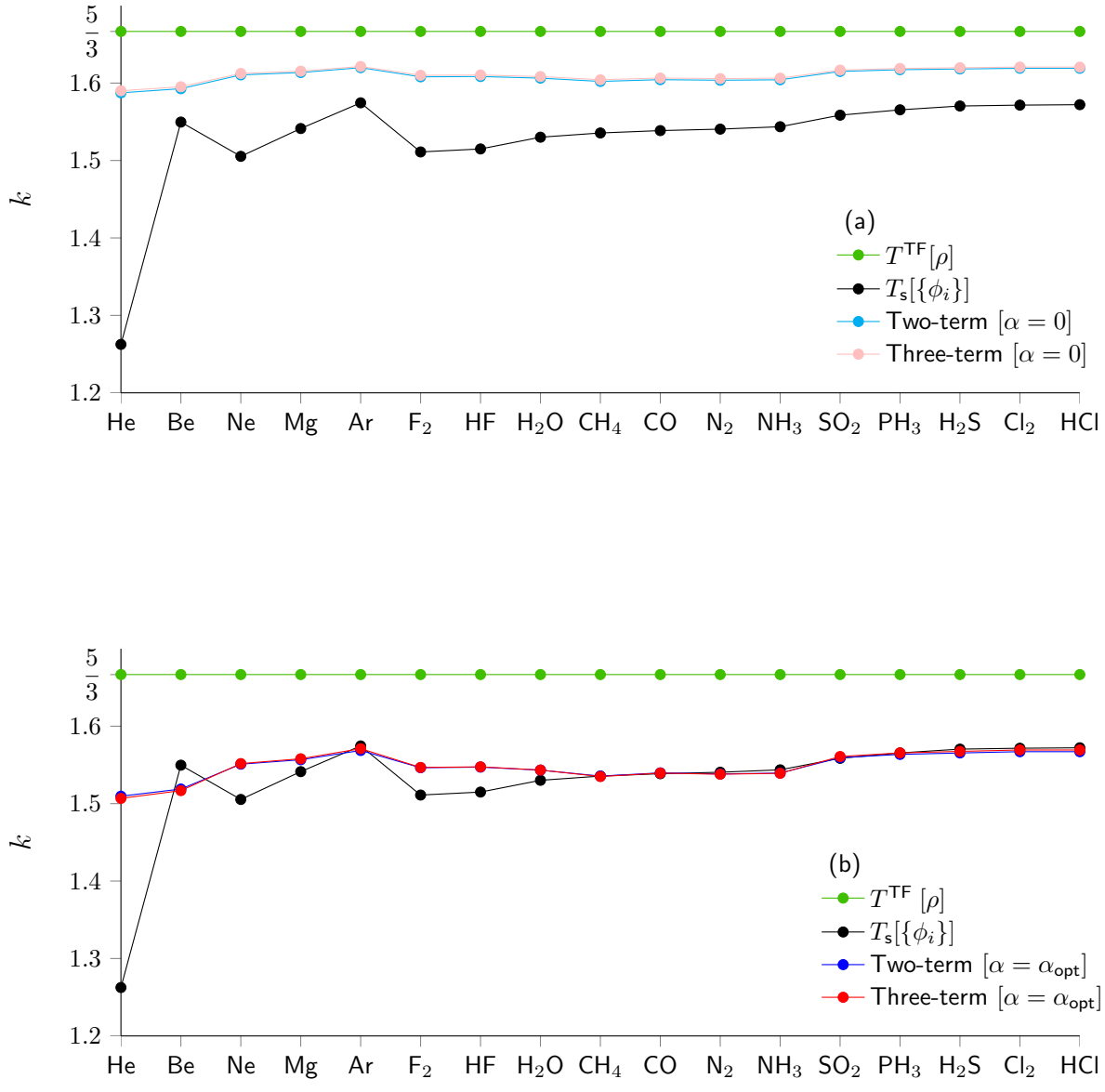

Figure 5: k values from two- and three-term functionals and the Thomas-Fermi functional, compared to near-exact values from the supplementary material of Ref. 26

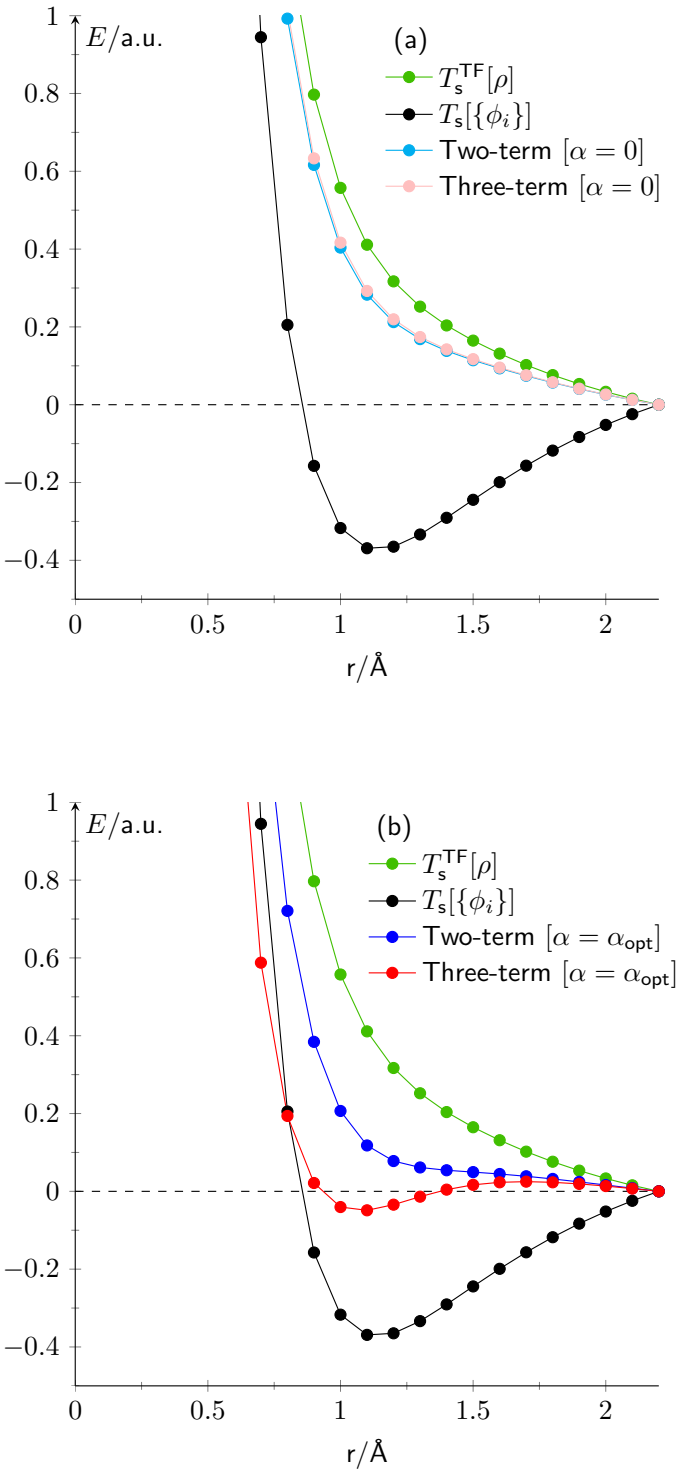

Figure 6: Potential energy curves of CO determined using two- and three-term functionals and the Thomas-Fermi functional, compared to the Kohn-Sham curve.

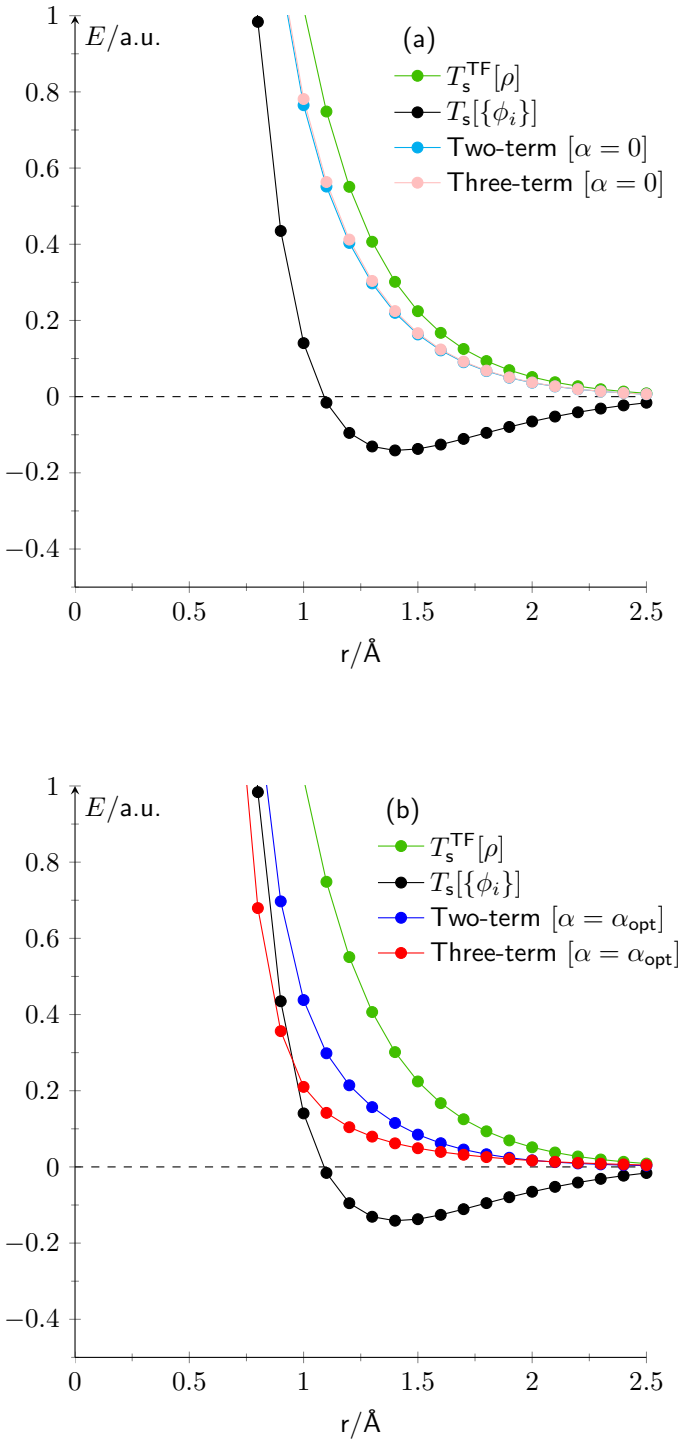

Figure 7: Potential energy curves of F_2 determined using two- and three-term functionals and the Thomas-Fermi functional, compared to the Kohn-Sham curve.

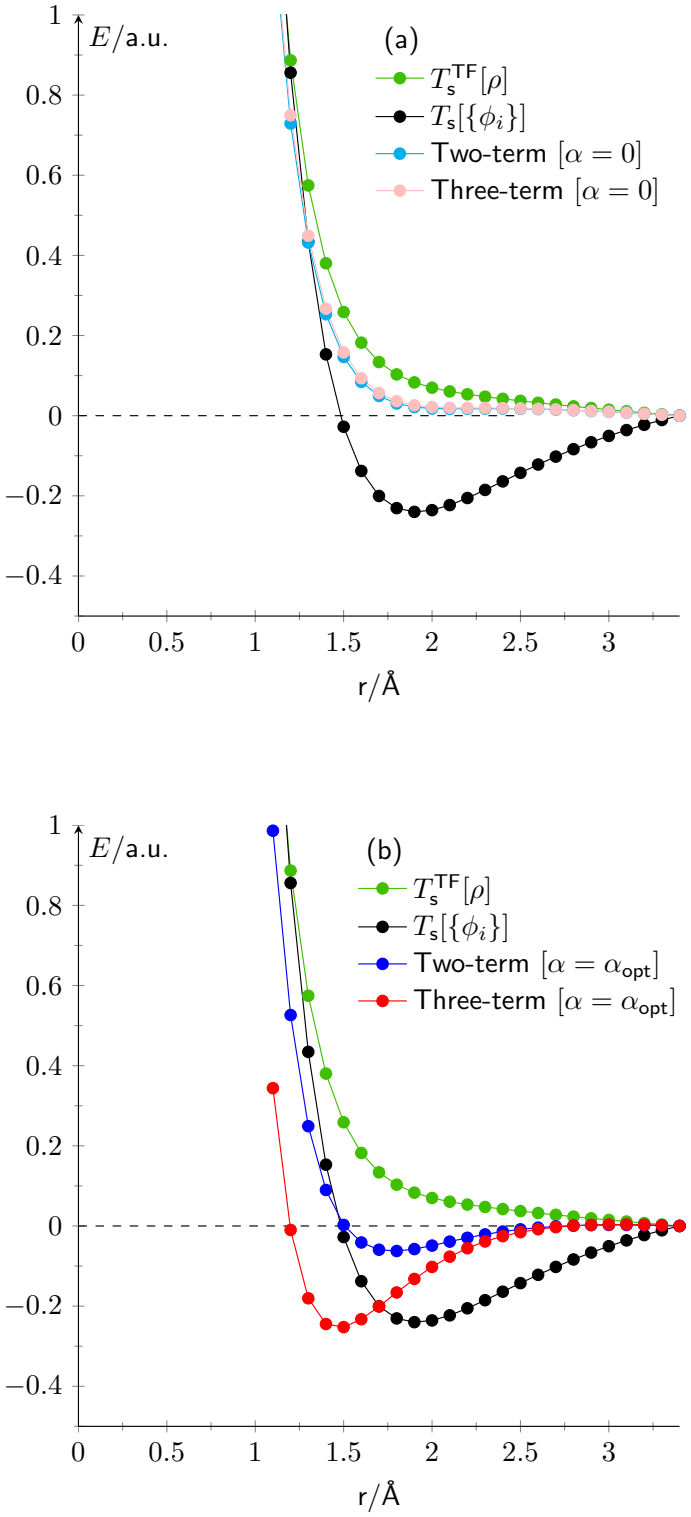

Figure 8: Potential energy curves of P_2 determined using two- and three-term functionals and the Thomas-Fermi functional, compared to the Kohn-Sham curve.